

NASA Technical Memorandum 87820

Problems in Merging Earth Sensing Satellite Data Sets

Paul H. Smith
Michael J. Goldberg

(NASA-TM-87820) PROBLEMS IN MERGING EARTH
SENSING SATELLITE DATA SETS (NASA) 15 p
Avail: NTIS HC A02/MF A01 CSCL 12B

N87-22457

H1/66 Unclass
0072167

MARCH 1987

NASA

NASA Technical Memorandum 87820

Problems in Merging Earth Sensing Satellite Data Sets

Paul H. Smith

Michael J. Goldberg

National Space Science Data Center

Goddard Space Flight Center



National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1987

FOREWORD

Satellite remote sensing systems provide a tremendous source of data flow to the Earth science community. These systems provide scientists with data of types and on a scale previously unattainable. Looking forward to the capabilities of Space Station and the Earth Observing System (EOS), the full realization of the potential of satellite remote sensing will be handicapped by inadequate information systems. There is a growing emphasis in Earth science research to ask questions which are multidisciplinary in nature and global in scale. Many of these research projects emphasize the interactions of the land surface, the atmosphere, and the oceans through various physical mechanisms. Conducting this research requires large and complex data sets and teams of multidisciplinary scientists, often working at remote locations. A review of the problems of merging these large volumes of data into spatially referenced and manageable data sets is presented in this paper.

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
1. INTRODUCTION	1
2. SITUATION	1
3. PROBLEMS, APPROACH AND STATUS	1
3.1 Data Management	3
3.2 Data Dissemination	3
3.3 Multisource Registration	5
3.4 Data Merging	5
4. THE NEXT GENERATION	5
4.1 Recommendations for Future Systems	5
4.2 Pilot Development	7
4.3 Concerns	7
5. FUTURE/SOLUTIONS	7
ACKNOWLEDGEMENTS	9
BIBLIOGRAPHY	10

PRECEDING PAGE BLANK NOT FILMED

1. INTRODUCTION

Satellite remote sensing systems provide a tremendous source of data flow to the Earth science community. These systems can provide scientists with data of types and on scales previously unattainable. Because of these opportunities, there is a growing emphasis in Earth science research to ask questions which are multidisciplinary in nature and global in scale. Many of these research projects emphasize the physical mechanisms of interactions between the land surface, the atmosphere, and the oceans. Conducting these research endeavors requires large and complex data sets and teams of multidisciplinary scientists, often working at widely dispersed locations. Unless adequate data and information systems are developed and employed, the full realization of the potential of satellite remote sensing, such as that expected through the Earth Observing System, Space Station and others, will be severely handicapped. In this paper, an overview is given of the problems expected to persist, at least into the near future, in merging these large volumes of Earth sensing satellite data into spatially referenced and manageable data sets.

2. SITUATION

The data and information situation in the remote sensing scientific community centers around three fundamental aspects that can be fairly simply stated. The situation is that there are (1) huge volumes of data, (2) many diverse data sources, and (3) a globally distributed and heterogeneous user community.

The first of the three aspects to consider is the volume of the data that must be handled. Present remote sensing systems such as NOAA's Advanced Very High Resolution Radiometer (AVHRR), the NASA-developed Thematic Mapper (TM), The French SPOT and others have already generated roughly a hundred terabits (i.e., 100×10^{12} bits) of sensor data. The sensors of the 1990's will be even larger suppliers of data with the Aircraft Imaging Spectrometer (AIS), the Synthetic Aperture Radar (SAR), the Moderate-Resolution Imaging Spectrometer (MODIS), and other sensor systems in the Space Station era promising to provide five times the total cumulative amount of remote sensed data (i.e., 500×10^{12} bits) every year!

Secondly, there are many satellite sensors or sensor systems which are and will be employed. These use different imaging technologies, have different spectral bands of various resolutions, and often present calibration difficulties and differences. The production of usable data products from these diverse sensors is receiving increasing attention and care. However, the data groups responsible for the products are becoming globally distributed and often are functioning with specific and unique purposes. The multitude of data formats continues to be a vexing problem.

Thirdly, not only are the sources of these remotely sensed data globally distributed with the user now needing to merge, in some appropriate fashion, data from more than one of the sources, but the users themselves are distributed around the world. The user communities are different. Four such user-group types are government, university, big industry, and small commercial. Even within a given type of user community, multi-purpose uses can be found, such as monitoring, mapping, specific scientific analysis or global change research studies. In addition, a given user or user institution will have widely varying capabilities for *using* the data, e.g. computers, software, expertise, time, money.

As is illustrated by the puzzled user in Figure 1, questions such as those posed are often not satisfactorily answered.

3. PROBLEMS, APPROACH AND STATUS

The situation described in Section 2 describes broadly-based and technically-deep problems that are beyond the scope of any single paper. However, it can be generally stated that Earth scientists do not want to overlook an

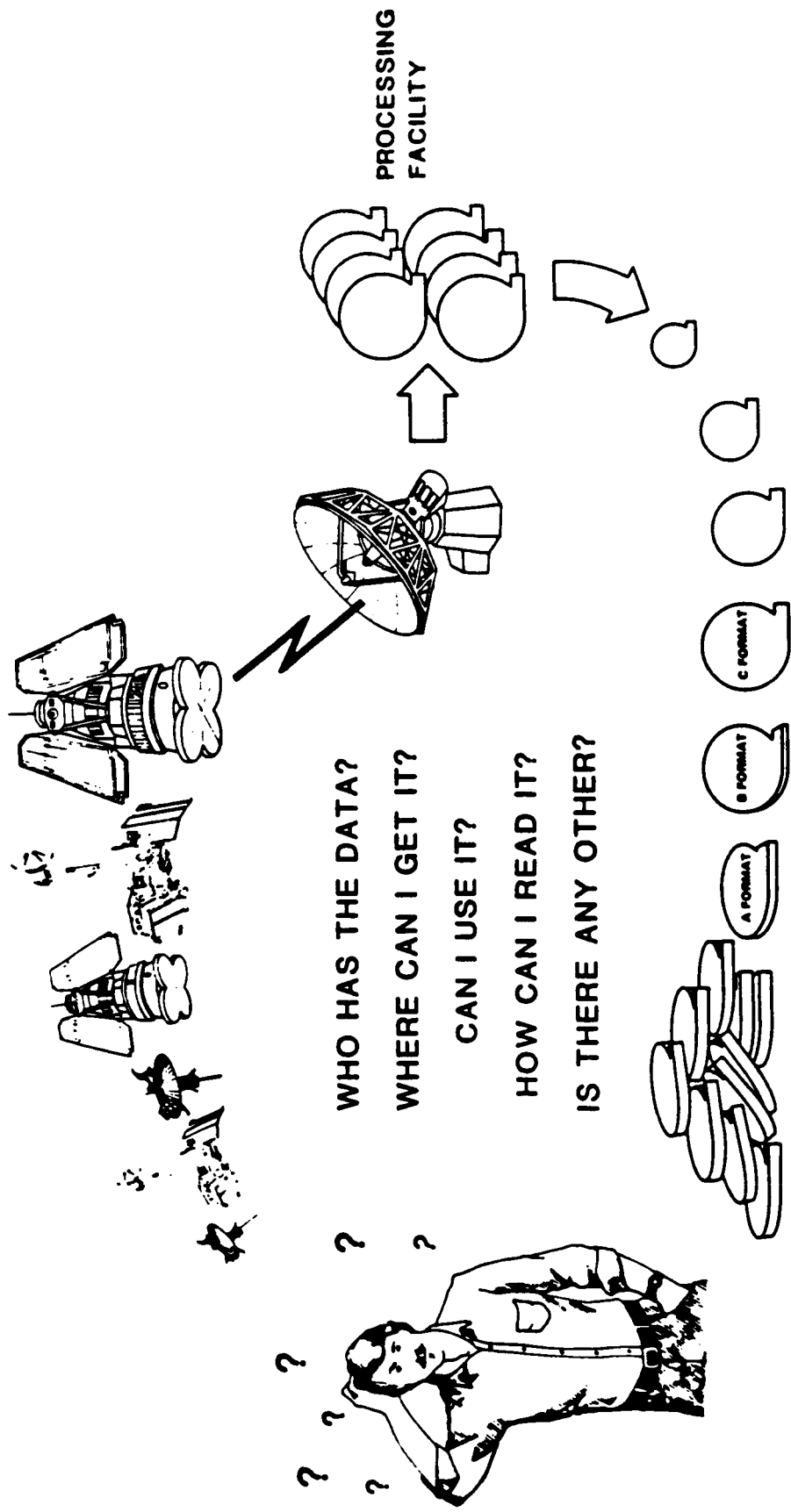


Figure 1.

important data set, wait a long time for data delivery, or spend a large proportion of their time and money converting data into useable form. To minimize these problems, current systems for the management, dissemination, and registration of multisource satellite and non-satellite data must be improved. This section will focus on the three sets of problems, (data management, data dissemination, and multisource registration), shown schematically in Figure 2, highlight key problems in each set, describe some approaches being taken to address the key problems, and give a brief status of those approaches.

3.1 Data Management

Data management systems for remote sensing basically organize satellite data, (and information about those data), in such a manner that scientists can rapidly determine what data are available for their particular research needs. Most present remote sensing data management systems do not support a wide enough variety of sensors or a flexible enough range of queries to satisfy the needs of interdisciplinary scientists. Historically, the earliest of these systems were set up only to manage data archives for particular satellites or sets of sensors. Each system gave access to a limited subset of remotely sensed data and provided no clues to the contents or even the existence of other related databases. Users had to have an intimate knowledge of specific sensor data characteristics in order to query the database and understand the system responses. More recent systems contain information about large groups of sensors related to a particular area of study, such as the atmosphere, land surface, or oceans. These systems provide on-line information about sensor data characteristics and provide a reasonably user-friendly and uniform method for querying the database about specific data holdings, regardless of the sensor. But the data holdings of these systems still do not support scientific investigations of land/atmosphere, ocean/atmosphere, and other interactions which cross discipline boundaries. In addition, several types of database queries which correspond to questions that scientists frequently ask about data are not generally available. Two of the most common questions relate to the availability of data for a specified portion of the Earth's surface and the quality (cloud cover, etc.) of such data. Even these queries are not supported for all sensor data types. The basic approach that is being taken toward the development of interdisciplinary Earth science data management systems is one of enhancement and integration of existing systems. The major enhancement that is being proposed is the development of a greater range of database queries. In particular, queries by latitude and longitude coordinates which describe rectangles or other closed boundaries are being developed so that multisource data can be identified for a given study site. The basic integration methodology that is being proposed is the development of automated directories that point to other database systems. This will at least provide the users of a given system with information on the existence, availability, and operation of other data management systems that may be of interest.

3.2 Data Dissemination

Earth science data dissemination systems are responsible for providing remote users with scientific data sets or with information about these data sets. Thus, these dissemination systems provide communications with data management systems as well as transmission or delivery of copies of selected data sets from their storage facilities to the scientific users. Current dissemination systems for geophysical data do not provide scientists with fast enough access to data management systems or to each others' data. While scientific data networks, telecommunications links, and mail facilities have increasing capabilities for transmitting data quickly over long distances, logistical problems in providing remote data management service and in establishing data interchange standards remain. Data management service is difficult to set up over distributed locations because of questions about payment for service, (particularly in the loosely structured scientific community), responsibility for database administration, and responsibility for day-to-day maintenance and operations. Meanwhile, data interchange is hampered by wide differences in hardware/software configurations, communications protocols, and user applications at various network nodes.

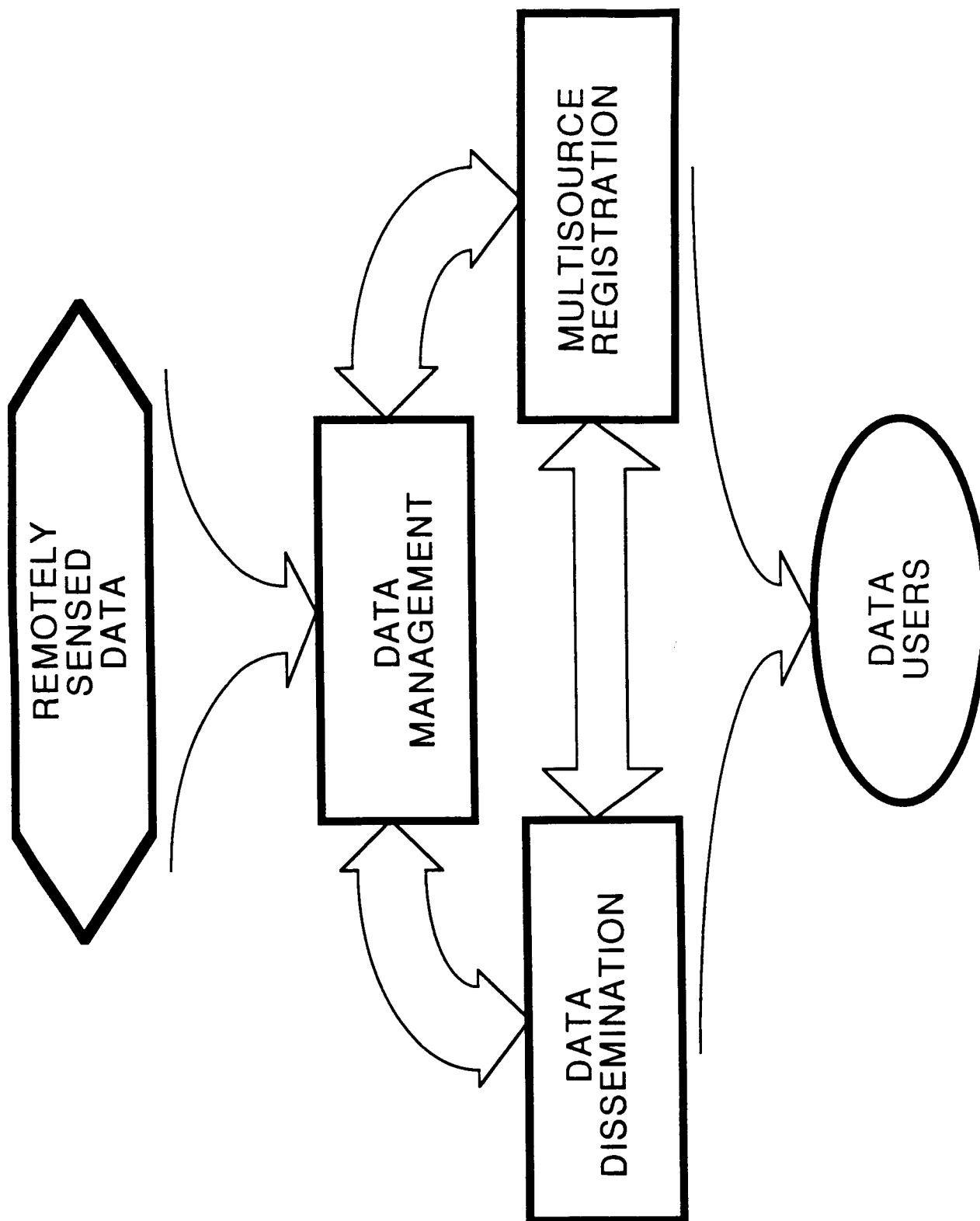


Figure 2.

Slow but steady progress is being made toward the resolution of data dissemination issues. In response to stated Earth science goals, institutions responsible for the management of very large geophysical databases, (e.g. NASA, NOAA, USGS, CCRS, etc.), are beginning to set up policies and procedures for interdisciplinary data exchange. Spatial data interchange standards adapted from the Landsat family of data types have been proposed. These standards combine the "common format family" and the "data definition language" approaches to data representation.

3.3 Multisource Registration

Multisource registration is the process of converting spatial (e.g. image and map) data sets into a common structure, geographic orientation, areal coverage, and resolution to enable easy intercomparison. The technology for the registration of multisource, interdisciplinary data is more advanced than that for data management or data dissemination, but scientists still spend too large a percentage of their time and money getting their data sets into usable form. As is the case with data dissemination, the technical advances, (in this case high-speed, high-volume image processing systems and sophisticated digitizers), are offset by logistical problems. Two major cost-producers are the needless duplication of effort incurred in writing and running image reformatting software and the labor involved in correcting minor inconsistencies in digitized maps. Because remotely sensed images are archived in such a wide variety of formats, scientists usually bear the cost of writing and running software for converting them into a form that is compatible with the local image processing system. Meanwhile, the correction of digitized maps is still highly subject to human error.

Progress is being made on both of the above mentioned problems. The cost of image data reformatting can be reduced in two ways. First, development of a generic image display format could be developed in conjunction with data interchange standards. Second, archive-to-image reformatting could be performed as a service by major data suppliers. A promising approach to the automatic editing and digitization of maps involves the use of knowledge engineering using expert systems technologies to duplicate the procedures used by experts to detect and correct errors.

3.4 Data Merging

While data merging is not an easy task, high quality products of merged data can be found. One such example is the Data Mosaic of the Western United States from AVHRR sensors during May, 1984 shown in Figure 3. This product is an Albers Equal Area Projection in color and is courtesy of R. J. Thompson, United States Geological Survey, EROS Data Center in Sioux Falls, South Dakota.

4. THE NEXT GENERATION

4.1 Recommendations for Future Systems

In the past few years, a number of studies, reports, workshops, and pilot systems have focused on the issue of improving handling and management of scientific satellite data. Mostly recently these activities are addressing the enormous difficulties referred to in Section 2.

One such activity is the recently published EOS Data Panel Report which has been recently published. Mike Ward, EOS Ground Systems Study Manager, NASA/GSFC, has kindly provided a summary view of the recommendations that have resulted from the EOS Data Panel's studies. The EOS Recommendations are for:

- Geographically distributed system
- Sites of varying capabilities and responsibilities

WESTERN UNITED STATES

AVHRR MOSAIC

0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300 1350 1400 1450 1500 1550 1600 1650 1700 1750 1800 1850 1900 1950 2000 2050 2100 2150 2200 2250 2300 2350 2400 2450 2500 2550 2600 2650 2700 2750 2800 2850 2900 2950 3000 3050 3100 3150 3200 3250 3300 3350 3400 3450 3500 3550 3600 3650 3700 3750 3800 3850 3900 3950 4000 4050 4100 4150 4200 4250 4300 4350 4400 4450 4500 4550 4600 4650 4700 4750 4800 4850 4900 4950 5000 5050 5100 5150 5200 5250 5300 5350 5400 5450 5500 5550 5600 5650 5700 5750 5800 5850 5900 5950 6000 6050 6100 6150 6200 6250 6300 6350 6400 6450 6500 6550 6600 6650 6700 6750 6800 6850 6900 6950 7000 7050 7100 7150 7200 7250 7300 7350 7400 7450 7500 7550 7600 7650 7700 7750 7800 7850 7900 7950 8000 8050 8100 8150 8200 8250 8300 8350 8400 8450 8500 8550 8600 8650 8700 8750 8800 8850 8900 8950 9000 9050 9100 9150 9200 9250 9300 9350 9400 9450 9500 9550 9600 9650 9700 9750 9800 9850 9900 9950 10000 10050 10100 10150 10200 10250 10300 10350 10400 10450 10500 10550 10600 10650 10700 10750 10800 10850 10900 10950 11000 11050 11100 11150 11200 11250 11300 11350 11400 11450 11500 11550 11600 11650 11700 11750 11800 11850 11900 11950 12000 12050 12100 12150 12200 12250 12300 12350 12400 12450 12500 12550 12600 12650 12700 12750 12800 12850 12900 12950 13000 13050 13100 13150 13200 13250 13300 13350 13400 13450 13500 13550 13600 13650 13700 13750 13800 13850 13900 13950 14000 14050 14100 14150 14200 14250 14300 14350 14400 14450 14500 14550 14600 14650 14700 14750 14800 14850 14900 14950 15000 15050 15100 15150 15200 15250 15300 15350 15400 15450 15500 15550 15600 15650 15700 15750 15800 15850 15900 15950 16000 16050 16100 16150 16200 16250 16300 16350 16400 16450 16500 16550 16600 16650 16700 16750 16800 16850 16900 16950 17000 17050 17100 17150 17200 17250 17300 17350 17400 17450 17500 17550 17600 17650 17700 17750 17800 17850 17900 17950 18000 18050 18100 18150 18200 18250 18300 18350 18400 18450 18500 18550 18600 18650 18700 18750 18800 18850 18900 18950 19000 19050 19100 19150 19200 19250 19300 19350 19400 19450 19500 19550 19600 19650 19700 19750 19800 19850 19900 19950 20000 20050 20100 20150 20200 20250 20300 20350 20400 20450 20500 20550 20600 20650 20700 20750 20800 20850 20900 20950 21000 21050 21100 21150 21200 21250 21300 21350 21400 21450 21500 21550 21600 21650 21700 21750 21800 21850 21900 21950 22000 22050 22100 22150 22200 22250 22300 22350 22400 22450 22500 22550 22600 22650 22700 22750 22800 22850 22900 22950 23000 23050 23100 23150 23200 23250 23300 23350 23400 23450 23500 23550 23600 23650 23700 23750 23800 23850 23900 23950 24000 24050 24100 24150 24200 24250 24300 24350 24400 24450 24500 24550 24600 24650 24700 24750 24800 24850 24900 24950 25000 25050 25100 25150 25200 25250 25300 25350 25400 25450 25500 25550 25600 25650 25700 25750 25800 25850 25900 25950 26000 26050 26100 26150 26200 26250 26300 26350 26400 26450 26500 26550 26600 26650 26700 26750 26800 26850 26900 26950 27000 27050 27100 27150 27200 27250 27300 27350 27400 27450 27500 27550 27600 27650 27700 27750 27800 27850 27900 27950 28000 28050 28100 28150 28200 28250 28300 28350 28400 28450 28500 28550 28600 28650 28700 28750 28800 28850 28900 28950 29000 29050 29100 29150 29200 29250 29300 29350 29400 29450 29500 29550 29600 29650 29700 29750 29800 29850 29900 29950 30000 30050 30100 30150 30200 30250 30300 30350 30400 30450 30500 30550 30600 30650 30700 30750 30800 30850 30900 30950 31000 31050 31100 31150 31200 31250 31300 31350 31400 31450 31500 31550 31600 31650 31700 31750 31800 31850 31900 31950 32000 32050 32100 32150 32200 32250 32300 32350 32400 32450 32500 32550 32600 32650 32700 32750 32800 32850 32900 32950 33000 33050 33100 33150 33200 33250 33300 33350 33400 33450 33500 33550 33600 33650 33700 33750 33800 33850 33900 33950 34000 34050 34100 34150 34200 34250 34300 34350 34400 34450 34500 34550 34600 34650 34700 34750 34800 34850 34900 34950 35000 35050 35100 35150 35200 35250 35300 35350 35400 35450 35500 35550 35600 35650 35700 35750 3

6

- Flexible, modular, layered architecture
 - expandable, evolvable system
- Centralized coordination and information management
- Distributed data archives
- Telescience - remote and interactive electronic access
- Research, operational, and commercial needs
- Transparent access to other NASA and non-NASA data sets
- Timely access to the science data
 - real-time data
 - archival data
- Data standards which allow data autonomy

The reader is referred to the full report of the EOS Data Panel for a complete discussion of the Panel's recommendations.

4.2 Pilot Development

NASA has just begun to build the Pilot Land Data System (PLDS). This is expected to be one of the most advanced data and information systems in the world for providing integrated data management, networking and communications, system access capability, land analysis software and special processes.

PLDS will include a central on-line directory of remote databases and data archives, distributed data catalog system, full information query capability about the data sets, state-of-the-art spatial data management functions, high speed communication lines servicing over thirty scientific investigator nodes by 1987, uniform access to a variety of NASA computer systems, use of both super computers and powerful workstations, library of software tools, comprehensive image analysis software, and advanced processing techniques and storage devices.

The functional concept of the distributed PLDS is shown in Figure 4. This illustrates distributed nodes at a few of the organizations expected to be participating with PLDS, with the user at any one of the nodes being able to access information and data at different places in the network to use various software tools for processing and analysis, and to perform the computations on the appropriate computer systems whether it be workstations or super computers.

4.3 Concerns

While there are many technical and scientific obstacles to overcome and great expectation of succeeding, some concerns remain which organizations such as CODATA and other international scientific groups might wish to consider. These are not easily answered or simply solved. Some of the concerns that come to mind are these: institutional problems that result in system incompatibilities; conflicting purposes and/or needs of remotely sensed data such as rapid use in forecasting, long-term studies of climatic change, detailed scientific investigations of physical mechanisms and studies of global change processes; data quality assessments that are at times non-existent or, at best, non-uniform; data archived at inappropriate levels of processing such that subsequent merging is difficult if not impossible, and cost effectiveness. Are the benefits in being able to merge remotely sensed data worth the cost required to solve the various difficulties?

5. FUTURE/SOLUTIONS

In summary, the problems of merging large volume data into spatially referenced and manageable data sets are formidable, but the future holds solutions to these difficulties and tremendous benefits can be derived. These

ORIGINAL PAGE IS
OF POOR QUALITY

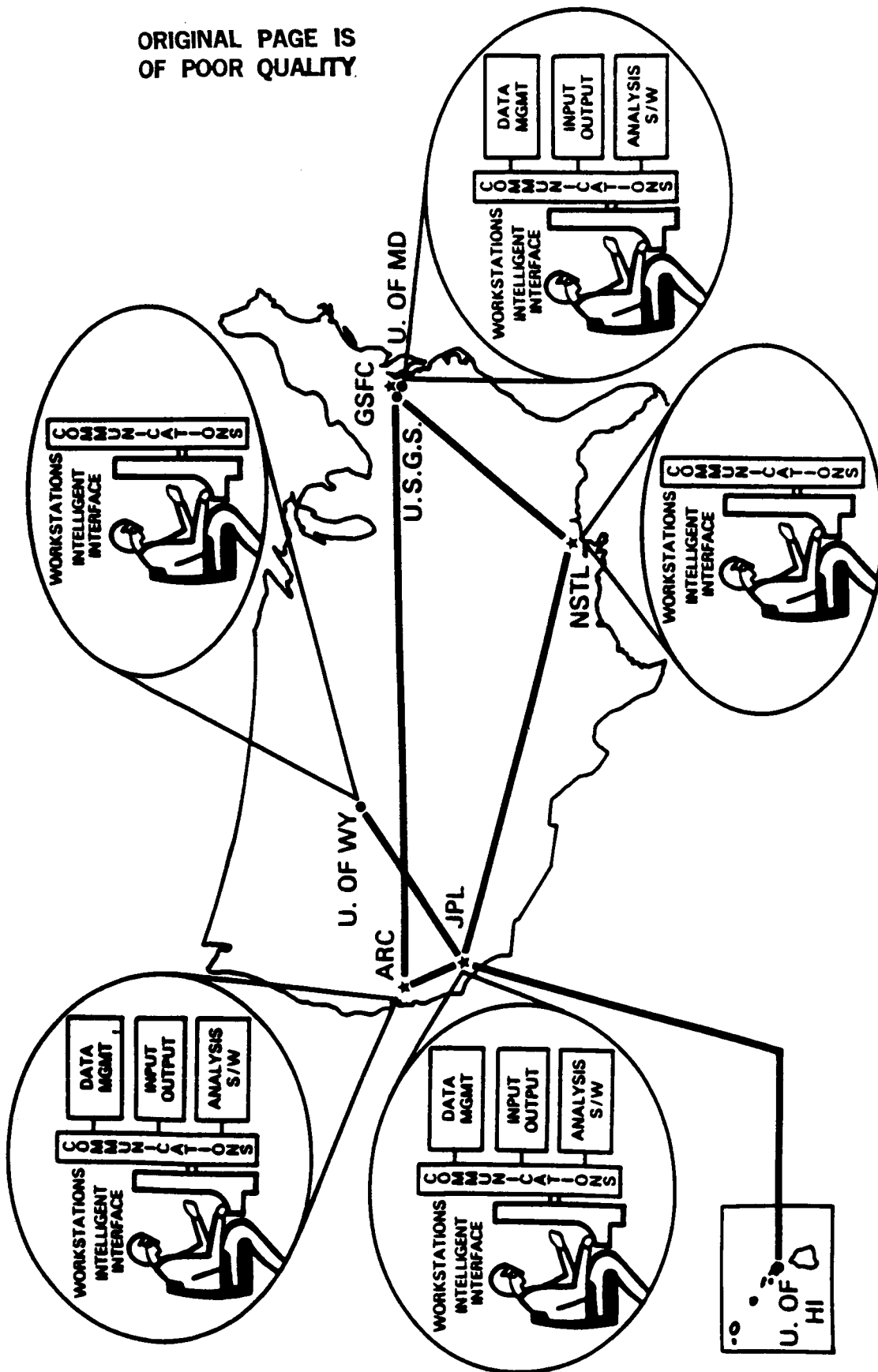


Figure 4. PLDS Distributed Functional Concept

difficulties can be overcome through the definition of common expectations, establishment of multidisciplinary teams, adoption of integrated system design and standard, development of distributed systems and the application of advanced technology.

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions toward the preparation of this review-style paper made by the following people:

David M. McKeown, Jr.

Carnegie-Mellon University

Donna J. Peuquet

Pennsylvania State University

Robert H. Rogers

Environmental Research Institute of Michigan

Philip H. Swain

Purdue University

A special note of appreciation goes to R. J. Thompson, USGS EROS Data Center, for providing significant materials for this paper. The contributions of Michael Ward and the authors of the papers cited in the Bibliography are also acknowledged.

BIBLIOGRAPHY

- Albert, T. M., "Geoscience Data Management," *Computers and Geosciences*, 9, pp. 3-6, 1983.
- Billingsley, F. C. and W. M. Strome, 'Standardization of Remote Sensing and GIS Data Transfer,' *ISPRS Commission II*, 1986.
- Campbell, William J., Paul H. Smith, Robert Price, Larry Roelofs, "Advancements in Land Science Data Management: Pilot Land Data System," *The Science of the Total Environment*, Elsevier Science Publications B.V., Amsterdam, 1985.
- Chase, R. P., Chairman EOS Data Panel, et al., "Report of the EOS Panel," *Earth Observing Systems Reports, Vol. Ila*, NASA TM87777, 1986.
- Estes, J. E., J. L. Star, J. G. Cressy, M. Devirian, "Pilot Land Data System," *Photogrammetric Engineering and Remote Sensing*, American Society of Photogrammetry, Vol. 51, No. 6, pp. 703-709, 1985.
- Goldberg, Morris, D. G. Goodenough, M. Alvo and G. M. Karam, "A Hierarchical Expert System for Updating Forestry Maps with Landsat Data," *Proceedings of the IEEE*, 73, 6, pp. 1054-1063, 1985.
- McKeown, D. M., Jr., "MAPS: The Organization of a Spatial Database System Using Imagery, Terrain, and Map Data," DARPA IUS Workshop, CMU-CS-83-136, 1983.
- Peuquet, D. J., "Data Structures for a Knowledge-Based Geographics Information System," Report for NASA, 1986.
- Peuquet, D. J., "The Use of Spatial Relationships to Aid Spatial Database Retrieval," *Proceedings 2nd IGU International Symposium on Spatial Data Sensing and Processing*, Seattle, July, 1986.
- Price, Robert D., Paul H. Smith, "NASA's Pilot Land Data System Development Program," *Peaceful Space and Global Problems of Mankind*, 36th Congress of the International Astronautical Federation, Pergamon Press, 1985.
- Rogers, R. H., personal communication, ERIM, 1986.
- Swain, P. H., J. A. Richards and T. Lee, "Multisource Data Analysis in Remote Sensing and Geographic Information Processing," *Symposium on Machine Processing of Remotely-Sensed Data*, 1985.
- Thompson, R. J., personal communication, USGS EDC, 1986.

1. Report No. NASA TM-87820		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Problems in Merging Earth Sensing Satellite Data Sets				5. Report Date March 1987	
				6. Performing Organization Code 634.0	
7. Author(s) Paul H. Smith and Michael J. Goldberg				8. Performing Organization Report No. 87B0275	
				10. Work Unit No.	
9. Performing Organization Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771				11. Contract or Grant No.	
				13. Type of Report and Period Covered Technical Memorandum	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546-0001				14. Sponsoring Agency Code	
15. Supplementary Notes Michael J. Goldberg is now employed by Computer Technology Associates, McLean, Virginia					
16. Abstract Satellite remote sensing systems provide a tremendous source of data flow to the Earth science community. These systems provide scientists with data of types and on a scale previously unattainable. Looking forward to the capabilities of Space Station and the Earth Observing System (EOS), the full realization of the potential of satellite remote sensing will be handicapped by inadequate information systems. There is a growing emphasis in Earth science research to ask questions which are multidisciplinary in nature and global in scale. Many of these research projects emphasize the interactions of the land surface, the atmosphere, and the oceans through various physical mechanisms. Conducting this research requires large and complex data sets and teams of multidisciplinary scientists, often working at remote locations. A review of the problems of merging these large volumes of data into spatially referenced and manageable data sets is presented in this paper.					
17. Key Words (Suggested by Author(s)) Satellite remote sensing systems, data sets, earth science research, data sources, user community			18. Distribution Statement Unclassified-Unlimited Subject Category: 66		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of pages 66	
				22. Price	